

# 周报

本周工作主要如下：

## 1.改进了上周计算模型参数方法

- 1）把一天分成 0:00-8:00、8:00-12:00、12:00-17:00、17:00-24:00 这四个时段，分别计算每天这四个时段下在家人数比例和外出比例
- 2）查询了每个城市每天四个时段对应的空气质量指数 AQI 和 pm2.5 值
- 3）由陈吟吟同学将结果代入回归模型中，计算变量系数，计算结果 pm2.5 的值为正，即 pm2.5 和外出人数比例与在家人数比例之差成正相关性。
- 4）根据何英华老师的建议，将一天分成 0:00-6:00、6:00-10:00、10:00-14:00、14:00-18:00、18:00-24:00 五个时段，分别按照前面 1-3 步重新计算和回归，结果仍显示 pm2.5 和外出人数比例与在家人数比例之差成正相关性。
- 5）如图 1 是第 4 步计算的湖州和嘉兴在家人数比例和外出比例及查询到的 pm2.5 和 aqi 数据值，其他四个城市的数据在此先省略。

城市	时间	在家人数	外出人数	总人数	在家人数/外出比例	min(AIQ)	avg(AIQ)	max(AIQ)	min(PM2.5)	avg(PM2.5)	max(PM2.5)	min(PM10)	avg(PM10)	max(PM10)
湖州	2013-12-17 06:00-10:00	319017	43287	362274	0.89596	0.119404	17	19	20	0	6	31	0	32
	2013-12-17 10:00-14:00	359515	43128	402643	0.892888	0.107112	20	25	31	31	35	40	32	37
	2013-12-17 14:00-18:00	363101	66569	429670	0.845069	0.154931	40	46	52	0	11	34	0	13
	2013-12-17 18:00-22:00	486903	131322	618225	0.787582	0.212418	52	52	52	0	0	0	0	0
	2013-12-17 22:00-2013-12-18 06:0	637865	89361	727226	0.877521	0.122879	52	52	52	0	0	0	0	0
	2013-12-18 06:00-10:00	632455	104448	736903	0.858261	0.141739	52	52	52	0	0	0	0	0
	2013-12-18 10:00-14:00	586483	160714	747197	0.784911	0.216089	56	62	68	47	52	57	70	75
	2013-12-18 14:00-18:00	566212	186136	752348	0.752593	0.247407	68	70	72	52	52	53	73	78
	2013-12-18 18:00-22:00	576251	175647	751898	0.766395	0.233605	72	76	84	52	74	115	75	99
	2013-12-18 22:00-2013-12-19 06:0	662112	103662	765774	0.864631	0.135369	84	100	114	115	135	147	133	149
	2013-12-19 06:00-10:00	624953	133080	758033	0.82444	0.17556	123	125	128	115	120	128	121	131
	2013-12-19 10:00-14:00	590767	171714	752481	0.771903	0.228197	128	143	150	120	126	131	139	154
	2013-12-19 14:00-18:00	574652	181089	755741	0.760382	0.239618	150	154	162	124	129	135	155	162
	2013-12-19 18:00-22:00	570107	181645	751752	0.758371	0.241629	162	170	179	135	154	169	164	188
	2013-12-19 22:00-2013-12-20 06:0	660766	104874	765640	0.863024	0.136976	179	189	203	163	181	198	210	223
	2013-12-20 06:00-10:00	629395	131363	760758	0.827326	0.172674	203	209	216	198	204	215	224	237
	2013-12-20 10:00-14:00	582649	171742	754391	0.772344	0.227656	216	221	227	183	202	219	219	243
	2013-12-20 14:00-18:00	569428	184266	753694	0.755516	0.244484	227	232	238	183	192	196	232	240
	2013-12-20 18:00-22:00	577274	171030	748304	0.771443	0.228597	238	241	244	195	202	215	233	247
	2013-12-20 22:00-2013-12-21 06:0	658839	106352	765191	0.861012	0.138988	232	238	245	82	119	151	107	153
嘉兴	2013-12-21 06:00-10:00	617981	136814	754795	0.81874	0.18126	226	229	232	144	152	169	176	185
	2013-12-21 10:00-14:00	560375	183446	743821	0.753373	0.246627	219	221	226	117	128	147	149	164
	2013-12-21 14:00-18:00	535234	199317	734551	0.728555	0.271345	217	218	219	120	138	163	147	162
	2013-12-21 18:00-22:00	595338	159033	718571	0.778682	0.221318	172	205	218	91	106	120	121	138
	2013-12-21 22:00-2013-12-22 06:0	613550	98144	711694	0.862098	0.137902	162	167	172	95	130	150	117	167
	2013-12-22 06:00-10:00	205167	71171	726338	0.742449	0.257551	168	168	171	136	140	146	151	160
	2013-12-22 10:00-14:00	193063	71180	264243	0.730627	0.269373	166	167	169	99	121	146	114	162
	2013-12-22 14:00-18:00	201248	51350	252598	0.796713	0.203287	154	156	161	81	86	92	94	102
	2013-12-22 18:00-22:00	176805	27320	204125	0.86616	0.13384	152	154	157	106	113	123	103	129
	2013-12-17 06:00-10:00	482473	67740	550213	0.876884	0.123116	31	31	33	27	28	30	29	33
	2013-12-17 10:00-14:00	595035	66858	661893	0.89899	0.10101	33	33	33	27	28	30	33	34
	2013-12-17 14:00-18:00	568335	152832	721167	0.788077	0.211923	33	33	34	18	22	28	21	26
	2013-12-17 18:00-22:00	604172	252451	856623	0.705295	0.294705	34	34	34	16	18	22	19	21
	2013-12-17 22:00-2013-12-18 06:0	886855	145059	1031914	0.859427	0.140573	34	36	40	22	30	38	19	39
	2013-12-18 06:00-10:00	880343	170930	1051273	0.837407	0.162593	40	42	45	35	40	49	61	63
	2013-12-18 10:00-14:00	797254	257603	1054857	0.755793	0.244207	48	50	52	43	44	46	68	70
	2013-12-18 14:00-18:00	751117	310753	1061870	0.707353	0.292647	52	54	56	44	57	80	68	82
	2013-12-18 18:00-22:00	756171	301874	1058045	0.714687	0.285313	56	63	72	80	91	106	101	120
	2013-12-18 22:00-2013-12-19 06:0	915762	155069	1070831	0.855188	0.144812	72	83	95	106	120	130	147	154
	2013-12-19 06:00-10:00	857660	208559	1066219	0.804394	0.195606	107	109	111	56	65	69	75	81
	2013-12-19 10:00-14:00	776863	281733	1058596	0.733862	0.266138	111	112	114	56	62	68	69	77
	2013-12-19 14:00-18:00	756004	305984	1061988	0.711876	0.288124	114	115	118	64	77	104	82	111
	2013-12-19 18:00-22:00	745049	311619	1056668	0.705093	0.294907	118	122	128	104	130	151	141	156
	2013-12-19 22:00-2013-12-20 06:0	908010	163711	1071721	0.847245	0.152755	128	137	150	151	170	185	170	206
嘉兴	2013-12-20 06:00-10:00	846659	221450	1068109	0.792671	0.207329	150	166	183	185	197	212	227	259
	2013-12-20 10:00-14:00	750929	308420	1059349	0.708859	0.291141	183	194	206	163	176	194	201	230
	2013-12-20 14:00-18:00	732441	324925	1057366	0.692703	0.307297	206	215	225	172	183	209	219	235
	2013-12-20 18:00-22:00	765967	292975	1058942	0.723332	0.276668	225	230	236	209	216	221	258	275
	2013-12-20 22:00-2013-12-21 06:0	901760	172077	1073837	0.839755	0.160245	217	229	238	69	102	157	87	145
	2013-12-21 06:00-10:00	838088	225554	1063642	0.787942	0.212058	190	204	217	69	76	82	87	99
	2013-12-21 10:00-14:00	734707	311827	1046534	0.702038	0.297962	173	181	190	82	90	98	98	118
	2013-12-21 14:00-18:00	701180	331083	1032263	0.679265	0.320735	150	162	173	90	95	106	133	138
	2013-12-21 18:00-22:00	746425	270713	1017138	0.733848	0.266152	133	142	150	97	109	123	124	141
	2013-12-21 22:00-2013-12-22 06:0	860650	136717	997367	0.862922	0.137078	119	121	123	80	97	109	95	123
	2013-12-22 06:00-10:00	224717	74232	298949	0.75169	0.24831	128	131	135	121	126	131	139	149
	2013-12-22 10:00-14:00	212519	79812	292331	0.726981	0.273019	133	134	135	65	88	125	83	121
	2013-12-22 14:00-18:00	221763	55187	276950	0.800733	0.199267	129	129	130	71	73	76	92	97
	2013-12-22 18:00-22:00	204466	32286	236752	0.863629	0.136371	126	127	128	97	109	120	128	137

图 1

- 6）以下是陈吟吟同学回归方法说明，摘自陈吟吟同学的回归说明文档。

## 天气与出行的回归分析

### 一、变量说明:

V1	V2	V3	V4	V5	V6	V7
$\ln(\text{外出人数/在家人数})$	当前时段 PM2.5(均值)	时 间 (周末 1, 工作日 0)	时 段 1(6:00-10:00)	时 段 2 ( 10:00-14:00)	时 段 3 ( 14:00-18:00)	时 段 4 ( 18:00-22:00)
V8	V9	V10	V11	V12	V13	
城市 1 ( 湖 州)	城市 2 ( 嘉 兴)	城市 3 ( 宁 波)	城市 4 ( 台 州)	城市 5(温州)	是否下雨 (1 是, 0 否)	

1. 因变量 v1,外出人数与在家人数之比对数
2. V4-V7 为时段的虚拟变量, 全 0 为 22: 00-次日 6:00
3. V8-V12 为城市的虚拟变量, 全 0 为舟山
4. V13 为是否下雨的虚拟变量, 由于天气数据以日为单位, 故当日数据认为是当日 6:00-次日 6:00

### 二、回归结果

用 V2, V3,V4,V5,V6,V7,V8,V9,V10,V11,V12,V13 对 V1 进行回归, 得到结果如下:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.5337340	0.1155919	-13.269	< 2e-16 ***
V2	0.0011781	0.0006271	1.879	0.062114 .
V3	0.0348897	0.0636217	0.548	0.584182
V4	0.0454274	0.0889988	0.510	0.610452
V5	0.0771942	0.0890293	0.867	0.387197
V6	0.0770146	0.0891097	0.864	0.388727
V7	0.0437003	0.0889817	0.491	0.624013
V8	-0.0498876	0.1151710	-0.433	0.665478
V9	0.0301291	0.1051312	0.287	0.774798
V10	0.0731698	0.1119701	0.653	0.514382
V11	-0.0321738	0.1109109	-0.290	0.772123
V12	0.0090042	0.1103055	0.082	0.935043
V13	-0.3412188	0.0986435	-3.459	0.000694 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3692 on 161 degrees of freedom

Multiple R-squared: 0.1904, Adjusted R-squared: 0.1301

F-statistic: 3.156 on 12 and 161 DF, p-value: 0.0004611

### 三、结果分析：

1. V2 的系数为正，说明 PM2.5 和外出的比例存在正相关，显著性为 0.062
2. V13 的系数为负，说明下雨会减少外出的比例，显著性接近 0。

### 四、R 语句

```
x=read.table("hazy weather.txt")
attach(x)
y=lm(V1~V2+V3+V4+V5+V6+V7+V8+V9+V10+V11+V12+V13);
summary(y)
```

### 2. 存在的问题

- 1) 小时为单位的 pm2.5 数据，只在<http://aqistudy.sinaapp.com/>该网站上查到过，但正如何老师在邮件中提到的，觉得在该网站上查到的部分数据有错误。但又找不到其他网站，做数据的对照。
- 2) 由于集群还不能访问，对之前林涛提供的轨迹数据是否正确的疑问还没有得到验证。